Journalism Studies

ISSN: 1461-670X (Print) 1469-9699 (Online) Journal homepage: www.tandfonline.com/journals/rjos20

Observing the Dynamics of the Online News Ecosystem

News diffusion processes among German news sites

Florian Buhl, Elisabeth Günther & Thorsten Quandt

To cite this article: Florian Buhl, Elisabeth Günther & Thorsten Quandt (2018) Observing the Dynamics of the Online News Ecosystem, Journalism Studies, 19:1, 79-104, DOI: 10.1080/1461670X.2016.1168711

To link to this article: https://doi.org/10.1080/1461670X.2016.1168711

Published online: 21 Apr 2016.

Submit your article to this journal

Article views: 1515

View related articles

View Crossmark data

Citing articles: 9 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rjos20

Check for updates

OBSERVING THE DYNAMICS OF THE ONLINE NEWS ECOSYSTEM
News diffusion processes among German news sites

Florian Buhl, **Elisabeth Günther**, and **Thorsten Quandt**

According to previous research, online news production is characterized by the volatility and fast pace of reporting as well as by the interplay of various contributors to news flows within the online news ecosystem. However, as both concepts are challenging to study, research discovering the dynamics of news flows at the ecosystem level is rare. Utilizing techniques of automated content analysis and big databases of online news content, this paper proposes an approach to make the pace and dynamics of news diffusion processes among online news sites observable. It reviews several lines of argument for the fast pace of online news production based on "breaking news" coverage, immediacy as a norm, and imitation among online news providers. Applying an analytical framework from diffusion of news events research, we find recurring dynamics for the diffusion processes of 95 events among 28 German online news sites. We distinguish three clusters of diffusion processes based on their temporal patterns. On average, it takes only 1.5 hours for a majority of online news sites to report news of the 43 events in the largest cluster. The narrow timing of broadly shared news decisions within the ecosystem reveals a strong potential for abrupt surges or bursts in online news flows.

KEYWORDS breaking news; immediacy; machine-based content analysis; news diffusion; news ecosystem; online news

Introduction

Two of journalism researchers' main foci in determining the peculiarities of online news production have been speed and the relatedness of news outlets in the online news ecosystem. The trend towards fluid deadlines of professional news outlets (Lewis, Cushion, and Thomas 2005; Lewis and Cushion 2009) and online journalists' publish-first-and-update-later routines (Karlsson and Strömbäck 2010; Saltzis 2012; Widholm 2016) have made the instantaneous release of information a defining attribute of the 24-hour news cycle (Rosenberg and Feldman 2008). Online journalists in particular consider immediacy to be a crucial success factor and a working imperative (Quandt et al. 2006; Boczkowski 2010). At the same time, the potentials of the networked public sphere (Benkler 2006) have opened professional news production to social media users in "network journalism" (Heinrich 2012), hence sparking renewed interest in the ecosystem of online news. Emerging news stories may circulate through various types of news outlets, e.g., from alternative media to legacy media to the blogosphere (Anderson 2010), thereby evolving by the contributions of both professional and citizen commentators

Journalism Studies, 2018
Vol. 19, No. 1, 79–104, https://doi.org/10.1080/1461670X.2016.1168711
© 2016 Informa UK Limited, trading as Taylor & Francis Group

(Im et al. 2011). Recently, journalism research has started to describe phenomena at the crossroads of both speed and the ecosystem perspective. For example, Hermida (2010) conceptualized the micro-blogging service Twitter as an always-on awareness system constituted by the contributions of its various users, and Spitzberg (2014) proposed a heuristic model of the diffusion of memes in new media.

However, empirical accounts observing the dynamics of the online news ecosystem are rare. To date, studies highlight either the notion of speed or the ecosystem-level of analysis. Karlsson and Strömbäck (2010) suggested three different approaches to analyzing the velocity of news items on single news sites. Studying the online newsroom of an Argentinian newspaper, Boczkowski (2010) demonstrated how strongly online journalists value the speed of publishing and how they routinely observe competitors' news sites. Widening the perspective to the ecosystem of online news, Weber and Monge (2011) described news flows among various types of online news providers, but without specifying their dynamics.

One crucial factor limiting our insights into the speed of online news flows is rooted in the difficulties of studying both processes and the ecosystem-level of analysis; the intersection of both makes the task even more challenging. On one hand, the analysis needs to account for the velocity of online news, the updating and re-writing of already published news items, sometimes their deletion. Hence, real-time recording of news content is required, which can be challenging even for single news sites (Karlsson and Strömbäck 2010; Saltzis 2012; Widholm 2016). On the other hand, data collection needs to be extended to a variety of online news sources simultaneously to map the ecosystem or subdomains of the ecosystem, at least. Obviously, the resulting multiple parallelisms of real-time recording of content can take efforts of human manual coding to their limits.

To overcome some of these problems, we propose an empirical approach that is based on a machine-based content analysis. We observe the dynamics of event-based news flows among professional online news sites as a subdomain of the ecosystem by developing an analytical framework that links the notion of speed in online journalism to diffusion of news events research (De Fleur 1987; Rogers 2000). We automatically downloaded and stored all news items published by 28 German online news sites during a 10-month time period between June 2013 and March 2014 (480,727 news articles in total). This database allows us to identify events relevant for analysis *ex post* and to relocate corresponding news items. Relating the points in time of the first news release on 95 events by all reporting online news sites, we can reconstruct how the timing of event-related publication decisions by each online news provider builds up to news flows at the ecosystem level. We focus our analysis of resulting news diffusion processes on two central characteristics of the potential speed of online news flows: (1) pace: how short is the time-span for a majority of online news sites within the ecosystem to report the same event?; (2) process dynamics: are online news flows characterized by recurring process patterns in terms of range, duration, and eruptions or bursts in the news flow? Contributing to the discussion of the prevalence of speed in online news production, the present study complements story-oriented accounts of the evolvement of news articles by single online news sites with an analysis at the media system level. By relating the timing of news releases of a large number of online news providers, we can reconstruct the dynamics of the resulting news diffusion processes and reveal common process patterns. The following literature review therefore focuses on the time dimension in online news production, with a special focus on technological opportunities and working routines among the online news providers.

Opportunities for Instantaneous "Breaking News" Coverage in Online News

Far before the rise of the internet, timeliness and journalists' orientation towards up-to-date events have played a role in news production (Galtung and Ruge 1965; Lewis and Cushion 2009). The efforts of news organizations to acquire and transmit timely information on a day-to-day basis date back to the diffusion of the telegraph at least (Risley 2000). While the production of newspapers is bound to daily deadlines, through radio and television, journalists have got faster ways at hand to disseminate the latest information. In the event of an extraordinary story, they can substitute their deadline-bound working routines with the production of "breaking news". Hence covering stories instantaneously, editors would interrupt regular program grids of radio and television channels (Berkowitz 1992). For example, on November 22, 1963, the news ecosystem responded to the attack on US President John F. Kennedy by rapidly disseminating the news from the spot through wire services and the majority of radio and television stations (Rosenberg and Feldman 2008, 18), so that, among those who heard of the shooting prior to the announcement that Kennedy had died, 70 percent were informed within 30 minutes (Greenberg 1964). In contrast to prototypical "news events" (Rogers 2000) like the Kennedy assassination or the 9/11 terrorist attacks, stories characterized by exceptionally high newsworthiness, the threshold for an event to be covered as "breaking news" has been decreasing since the rise of 24-hour news channels like CNN in the 1980s, as the pretension of timeliness is one of their unique selling propositions (Rosenberg and Feldman 2008; Lewis and Cushion 2009). Technological and organizational potentials to disseminate any information immediately have turned into an affordance of news production. Nowadays, it is not uncommon that 24-hour news channels present their audiences with "breaking news" stories that lack both unexpectedness and surprise. For example, on January 12, 2006, British Sky and BBC News 24 reported the announcement of the Bank of England that interest rates would remain unchanged as "breaking news", although both point in time and content of the announcement had been anticipated (Lewis and Cushion 2009, 310).

The expansion of the news market to the internet created the opportunity for news organizations publishing via text-based media like newspapers and magazines to enhance the timeliness of their information output, too. Newspaper editors were now able to reach their readers quickly and independently from deadlines of print editions through their online news sites (Risley 2000). When they become aware of dramatic situations from the "news event" type, the fluidity of deadlines online allows them to cover these stories instantaneously (Risley 2000; Rosenberg and Feldman 2008). Additionally, online communication tools have contributed to the availability of stories appropriate for "breaking news" coverage: the proliferation of portable devices for digitally recording events and transmitting news has increased the opportunities for real-time reporting, even of unexpected events (Lewis and Cushion 2009). The openness of the "network journalism" (Heinrich 2012) ecosystem for contributions from outside professional news organizations combined with the always-on awareness system of social media like Twitter (Hermida 2010) provide the editors of online news sites with a larger pool of current information to develop into news stories than ever before. As the technological opportunities both for prompt, deadline-unbound news coverage and for fast information gathering are ubiquitous in the online environment, we expect them to contribute to narrowing the time-frame of first

reports about extraordinary events by a variety of news sites and thus to generate brief news diffusion processes at the ecosystem level.

Immediacy as a Production Norm in Online Newsrooms

Similar to the expansion of "breaking news" on 24-hour television news channels, in online news, the instantaneous release of information is not limited to dramatic events whose relevance for a wide audience is unambiguous. Rather, the fluidity of deadlines has developed into a vanishing of deadlines (Karlsson and Strömbäck 2010). As a result, accelerated news cycles necessitate the provision of new information to disseminate to the audience and therefore extend instantaneous reporting beyond stories characterized by extraordinary newsworthiness (Rosenberg and Feldman 2008). This phenomenon has granted immediacy a "mythological status" (Lim 2012) in the scholarly discussion of online news production:

> The term immediacy refers to the notion that the news cycle with respect to online news … has become radically shortened (Singer 2003) and that the time lag between when a news organization becomes aware of an issue and publishes information about it has been radically shortened. (Karlsson and Strömbäck 2010, 4)

The technological potential for permanently updating online news sites appears to have turned the immediate release of information into a core principle of online news production (Quandt et al. 2006; Lim 2012). For example, for the newsroom of the *Washington Post*, editor Howard Kurtz reported: "In the last year, the pendulum has swung in our newsroom to putting things on the Web almost immediately, with the exception of some big exclusive story or long investigative piece" (Rosenberg and Feldman 2008, 139). Immediacy can be considered institutionalized as a working routine by a majority of online news sites (Boczkowski 2010) and seems to have partly displaced other criteria for the evaluation of the quality of news outlets (Lewis, Cushion, and Thomas 2005; Lewis and Cushion 2009; Saltzis 2012). Being among the quickest to publish a story has become a key strategy to prevail against competitors in the news market (Lim 2012). Online journalists both in the United States and in Germany perceive getting information to the public quickly as the most important dimension of their professional roles (Quandt et al. 2006). Moreover, they believe their readers expect immediacy from online news reporting (Boczkowski 2010). In sum, we assume that the role of immediacy as a working routine has prompted online newsrooms to widen the scope of stories to be released as soon as possible beyond dramatic events. Considering the fluidity of deadlines, "as soon as possible" is often almost equivalent to instantaneous publication.

It is important to note, however, that the prevalence of immediacy in online news is limited by other working routines (Lim 2012). While each online newsroom will be eager to cover "news events" instantaneously, we expect them to apply different publication strategies tailored to their specific audiences when it comes to less extraordinary events. For example, newsrooms of online news sites designed as news hubs for national audiences will feel stronger obligations to follow national politics in a very timely manner than online newsrooms of local newspapers. Therefore, we expect more variation among online news sites in the timing of publication for less extraordinary events, so that the overall time-span of respective news diffusion processes within the news-site ecosystem will be longer.

Orientation Towards Competitors Driven by the Pursuit of Immediacy

Editors of online news sites assume their readers evaluate the timeliness of their updates against the points in time of competing sites covering events (Boczkowski and de Santos 2007; Boczkowski 2010). That is why the immediacy of online news can also be "considered to be publishing speed for news producers, as compared with their competitors" (Lim 2012, 73). Therefore, online newsrooms have established a working routine of monitoring other news sites and adopting news stories in order not to lag behind (Boczkowski and de Santos 2007; Boczkowski 2010). In a survey, Quandt et al. (2006, 178) found that 79.4 percent of German online journalists and 88.0 percent of US online journalists "keep up with the news by reading the websites of other news organizations" on a daily basis. Doubtless journalists' orientation towards competing news organizations was part of their working routines well before the online era. Monitoring their competitors' publication activities helped journalists reduce uncertainties about their own news decisions (Donsbach 2004). Resulting imitation of issue orientation contributed to the convergence of attention across media documented in intermedia agenda-setting research (Reese and Danielian 1989; Vliegenthart and Walgrave 2008; Lim 2011).

Yet, monitoring and imitation seem to be motivated by different or at least additional ends in online news production, as they apply to publication decisions of events, whose newsworthiness appears unambiguous, too (Boczkowski and de Santos 2007; Boczkowski 2010). If the editors of an online news site cannot be the first to report an event, they will at least try to avoid lagging behind the publication times of competing sites in the market too much. Therefore, it seems reasonable for online newsrooms to establish a working routine of permanently observing those online news sites which have proved to be among the first to report emerging stories (Boczkowski 2010). Certainly, the implementation of this observation-and-imitation routine depends on how much each newsroom pursues immediacy of coverage of specific issues. The routine has been studied for competing news sites within very homogenous audience markets so far (Boczkowski and de Santos 2007; Boczkowski 2010). The smaller the overlap between the markets of two online news sites is, probably, the lesser the motivation in their newsrooms to orient towards each other will be. But, for example, following the coverage of national politics on online news sites functioning as authorities for these issues (Weber and Monge 2011) appears as a useful routine for many online newsrooms of local newspapers. Therefore, another rationale for the expectation that online news sites report an event within a short time-span from each other is that the pursuit of immediacy is today met by the technical infrastructure to observe the news production of competitors in real-time. At the ecosystem level, the close timing of reporting is mirrored in the brevity of news diffusion processes as a whole.

News Event Diffusion as Analytical Framework for the Ecosystem Perspective

To describe the dynamics of online news flows at the ecosystem-level of analysis, we embed the assumptions about the timeliness of online publishing into an analytical framework derived from news event diffusion research (De Fleur 1987; Rogers 2000). This research tradition studied the amount of time necessary for the population to become aware of major events through mass media and interpersonal communication. As the main interest was in the potential brevity of this time-span, the events chosen for

investigation were usually characterized by exceptionally high newsworthiness: "news events" (Rogers 2000) like the Kennedy assassination (Greenberg 1964) and the 9/11 terrorist attacks (Rogers and Seidel 2002). Researchers found the dynamics of news diffusion processes to be characterized by a distinct pattern of spurts, surges, or bursts: the combination of mass-media dissemination of information with subsequent word-of-mouth regularly resulted in S-shaped diffusion dynamics with low diffusion rates at the beginning and towards the saturation of processes and a burst midway through (Rogers 2000).

With the rising popularity of online communication, this research tradition has come to life again. On the one hand, online news sources and computer-mediated, distant interpersonal communication were expected to reshape news diffusion in society. However, the contribution of both mass-media and interpersonal channels of online communication to the diffusion of news about the 9/11 terrorist attacks (Rogers and Seidel 2002) and the Columbia shuttle breakup (Glascock and King 2007) was still low in the early 2000s. On the other hand, research turned to the question of how the diffusion of information within the networked news ecosystem is intertwined with processes of social production online (Benkler 2006). Anderson (2010) studied the emergence and development of a local news story in the online political press and in traditional newspapers. Im et al. (2011) retraced the evolvement of two news stories by diffusion through various arenas of online communication. In these studies, interest in the dynamics of diffusion processes was secondary at most.

In the present study, we apply the analytical framework of news diffusion to explore the process patterns of news flows among online news sites. We ask how much time it takes for an event to be reported by a large sample of online news sites, which represent the major subdomain of the online news ecosystem. More specifically, based on a variety of events, we aim to describe generalized patterns of news diffusion processes among online news sites in terms of their range, duration, and dynamics. As we are primarily interested in the potentially fast pace of diffusion processes, we focus the analysis on extraordinary "news events" and on events whose newsworthiness at least seems unambiguous enough be reported by a majority of news sites. We assume that the crucial determinants of the timing of publication, namely technological opportunities for "breaking news" coverage and the working routines of immediacy and imitation, to be—to varying degrees—similar among the variety of online news sites. Consequently, we expect close timing of their individual event-related publication decisions, which, at the ecosystem level, result in brief diffusion processes. The closer the timing of publication decisions among various online news sites becomes, the more likely is the observation of bursts as part of the dynamics of news flows.

> **RQ1:** What are the characteristics of news diffusion processes among online news sites with regard to range, duration, and temporal dynamics?

For these types of stories, we expect diffusion to generally proceed within short timespans, as "breaking news" coverage, pursuit of immediacy, and the imitation of first-reporting online news sites should apply for many online newsrooms. Especially, major "news events" should be reported instantaneously by a majority of online news sites, whose closely timed individual responses would result in diffusion processes exhibiting the dynamic pattern of bursts at the ecosystem level. This is even more so as we deliberately focus our analysis on extraordinary "news events" to be able to reconstruct diffusion processes with a high range of reporting news sites. On the other hand, we expect the

responses of the online newsrooms in our sample to be more spread out over time for events that are relevant enough to be reported by a majority of news providers, but are either less extraordinary or less easy to report (e.g., due to necessary background checks), so that respective news diffusion processes take more time and are less likely to exhibit bursts. We aim to reveal these differences in recurring patterns of news diffusion, depending on the homogeneity of publication timing across the various news sites:

> **RQ2:** Do news diffusion processes among online news sites exhibit common dynamic patterns across a variety of events?

Contrary to previous studies of online news flows, the reconstruction of diffusion processes within the online news ecosystem in this study is neither based on mentions of sources (Messner and DiStaso 2008) nor on hyperlinks explicitly referring to the same content (Weber and Monge 2011). Instead, reference to the same event is the crucial criterion for an online news site to be attributed to a specific news diffusion process. This content-based approach appears as a valuable complement to reference-based studies: the reconstruction of news diffusion processes in general and within the online news ecosystem in particular does not aim at direct relationships among the population under study, which may be inscribed into text or software traces, but more generally at process patterns, which emerge from the timing of both dependent and independent adoption decisions by individual elements (in the present study: online news sites). Although we introduced several potential explanations for the assumed brevity of news diffusion processes online, ranging from opportunities for "breaking news" coverage to imitation, we accordingly do not intend to make an empirical distinction between those underlying mechanisms. Rather, our primary goal is to lay the analytical and methodological foundations for a systematic description of the temporal patterns of news diffusion processes online. In order to achieve this goal, we base our study on a large-scale, automated data collection that is analyzed by means of a machine-based content analysis.

Method

Sample

Sampling unit. As a coherent subdomain of the online news ecosystem in terms of professional working practices, we chose a comprehensive sample of the online news sites of print newspapers in Germany. Despite similarities of their organizational settings, these online news providers target diverse audiences ranging from local to national. Variations of working routines therefore appear likely regarding their pursuit of immediacy when covering national politics beyond extraordinary events. The combination of similarities and variations of newsroom working routines within the German online news ecosystem should provide ample opportunities to observe both recurrences and systematic differences of news diffusion process patterns depending on the types of events. For an overall sample of German online news sites, we relied on the key data provided by Schütz (2012), listing all print newspapers published in Germany in 2012. Using a custom Web crawler, all print newspapers providing an online news site with an RSS feed focusing on the respective homepage were retained for analysis. Since RSS is a common online standard, this criterion did not noticeably influence our sample, but provided considerable benefits with regards to a reliable data collection. Our sample comprises 28 websites,

covering all major news outlets in Germany: *Abendzeitung München*, *Augsburger Allgemeine*, *Berliner Kurier*, *Berliner Morgenpost*, *Berliner Zeitung*, *Bild*, *B.Z. Berlin*, *Frankfurter Allgemeine Zeitung*, *Frankfurter Rundschau*, *Hamburger Abendblatt*, *Hamburger Morgenpost*, *Handelsblatt*, *Kölner Express*, *Kölner Stadtanzeiger*, *Leipziger Volkszeitung*, *Mitteldeutsche Zeitung*, *Neue Ruhr Zeitung*, *Potsdamer Neueste Nachrichten*, *Spiegel*, *Süddeutsche Zeitung*, *die tageszeitung* (*taz*), *tz München*, *Westdeutsche Allgemeine Zeitung* (*WAZ Niederrhein*), *Welt*, *Weserkurier*, *Westfalenpost*, *Westfälische Rundschau*, and *DIE ZEIT*.

A comprehensive data collection was carried out from June 11, 2013, to March 11, 2014: every hour, all articles published via the websites' RSS feeds were automatically downloaded and stored in a MySQL database. This way, we collected 480,727 online news articles. Their main textual content was extracted automatically using the boilerpipe library (Kohlschütter, Fankhauser, and Nejdl 2010): examining the number of words and hyperlinks per paragraph, this algorithm automatically evaluates which parts of an article belong to the main textual content, and which ones consist of HTML markup, advertorial content, or website navigation elements.

Unit of analysis: identification of event-related online news reports. We defined three criteria to identify relevant events that occurred during the sample period. (1) The event has to attract interest: as we aim to describe news diffusion processes, the event has to be newsworthy to a majority of online news sites in our sample. (2) The event has to be unexpected: for easily expected events, for example the inauguration of a previously elected politician, news coverage can start weeks ahead of the point of time it actually happens. In these cases, we cannot define a starting point for the news diffusion process, which makes it difficult to isolate reports on the event itself. Therefore, we restrict our analysis to events with a clear-cut starting point. (3) The event can be described unambiguously with few keywords: this is a technical prerequisite for the following step, in which we automatically searched the database for all reports on the events. To keep the effort in the selection phase low, we manually scanned every downloaded news article from two online news sites for each day of the time period under study to identify events meeting all three criteria. This procedure yielded a pool of 131 events within the sampled time-frame, whose diffusion processes among online news sites were to be reconstructed (see Figure 1).

First, we identified every online news article related to each of the 131 events by means of an event-based database query. For each event, we defined a regular expression with one or two mandatory keywords and up to two optional keywords: to be identified as event-related, an article has to contain keyword1 and keyword2 in combination with either keyword3 or keyword4. For example, in case of the birth of Prince William's and Kate Middleton's "Royal Baby" Prince George, "William" had to be mentioned in combination with either "birth" or "son". Search expressions regularly included the main actors involved in an event (names of persons, but also of institutions and organizations such as the National Security Agency (NSA), the Italian Senate, Boeing, and Microsoft), the scene (e.g., the island of Lampedusa, London Heathrow airport), or dense descriptions of the event if available (e.g., birth, train accident). The keywords for each event were developed and tested during a manual pretest, which required several manual refinements of search expressions (see Appendix A for the final list of keywords for all events, and Günther and Quandt [2016] for a more general description of this research process). The database query resulted in 61,132 articles that matched one of the expressions for the 131 events (see Figure 1). Second, we identified the very first online news article reporting each event, removed all

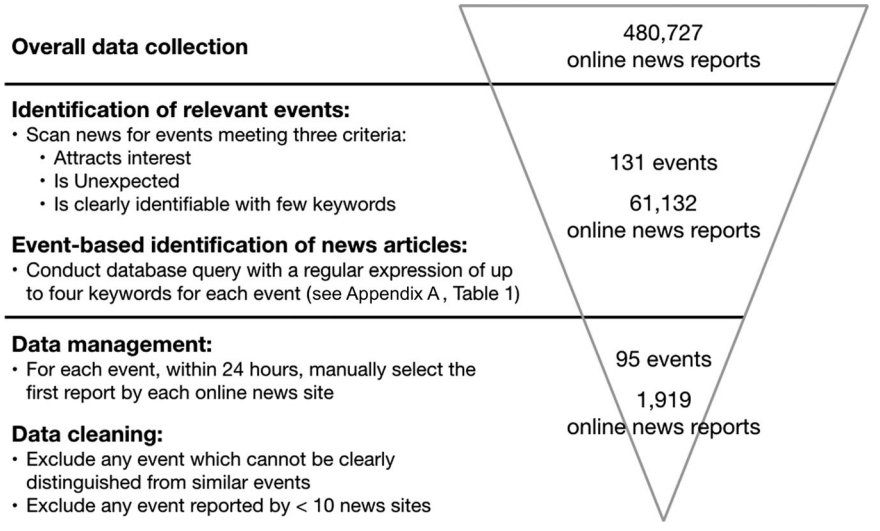


FIGURE 1
Stepwise procedure to identify relevant events and related online news reports

event-related articles that were not published within the following 24 hours, and manually selected the first report of each event by each online news site. For five cases, this procedure did not generate valid reconstructions of event-specific subsamples of news articles, as similar events were too close in time to be clearly distinguished from each other. We excluded these events from the analysis. Finally, to restrain the scope to events characterized by high newsworthiness, we determined a minimum range of the diffusion of news about the respective event, meaning that events with reports by fewer than 10 news sites were excluded. The final sample comprised 95 events that were reported in $N = 1919$ articles on 28 online news sites (see Figure 1).

Data analysis. With the help of automatically derived time stamps to the second for every online news article, we calculated the time lag between the first-reporting article of each online news site and the very first report on the respective event in the news-site ecosystem. The time-lag data allow us to relate the accumulated number of online news sites having reported an event to the amount of time elapsed since the event has been reported for the very first time. Therefore, we can observe the length of the full diffusion process for each event, i.e., the amount of time elapsed until the last online news site has reported the event within 24 hours, but we can also answer the question, how many online sites have reported the event within one, two, or three hours, for example. Drawing curves for the diffusion of each event, we can make this information visually available for any point in time during diffusion processes (see Figures 2 and 3). Moreover, the dynamics of diffusion processes, i.e., speed and range at which reporting online news sites join the diffusion process, are easily accessible through visual inspection of diffusion curves. To statistically separate homogenous subsamples of news diffusion processes with regard to their dynamics, we conducted a hierarchical agglomerative cluster analysis with the diffusion processes of each event as cases. We deliberately decided against controlling for the length of diffusion processes, as we consider the duration itself a relevant piece of information to identify

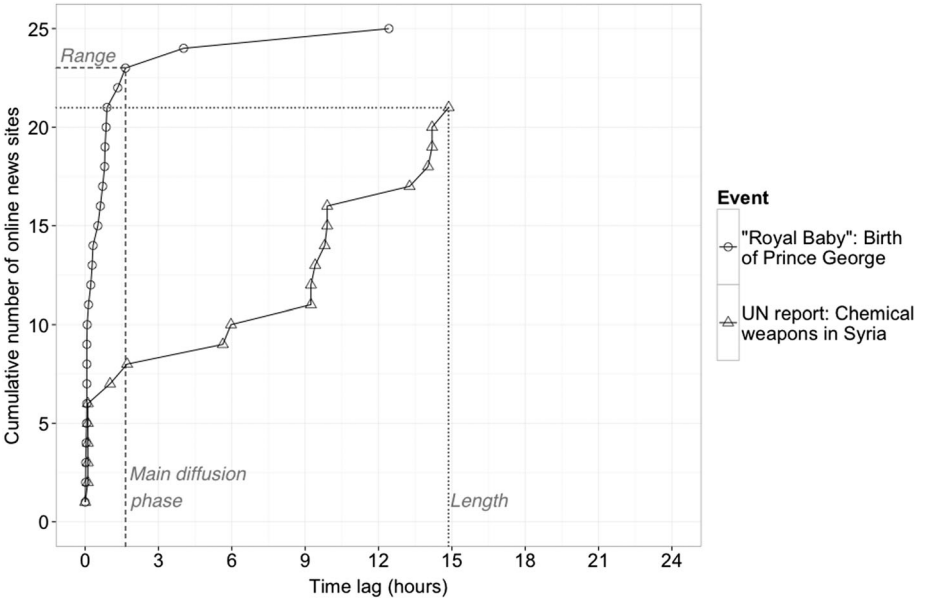


FIGURE 2
Diffusion processes of the "Royal Baby" case and the UN report on chemical weapons in Syria

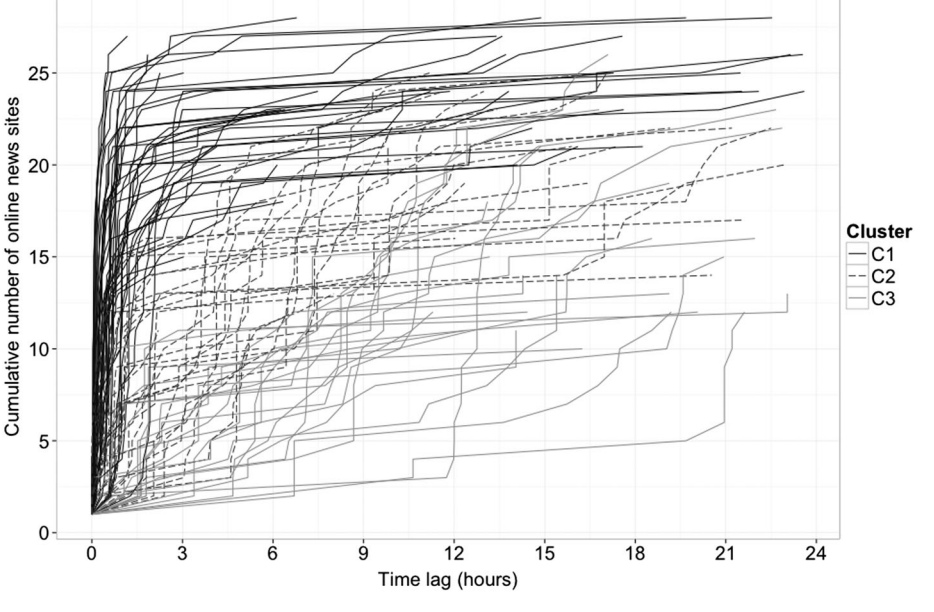


FIGURE 3
Diffusion processes of all 95 events, grouped by dynamics-based cluster membership

similar patterns of diffusion. The structure of the dendrogram was clear-cut, suggesting a three-cluster solution (see Appendix B). In a *post-hoc* analysis, we determined the types of events typical for each cluster. Finally, we also submitted the time-lag data to event history analysis according to Kaplan–Meier to summarize and compare subsamples of event-based news diffusion processes.

Results

Dynamics of News Diffusion Processes Among Online News Sites

For each of the 95 events, Figure 3 displays a curve of the corresponding news diffusion process resulting from the accumulated number of online sites reporting the event at any point within 24 hours. To illustrate better how we analyzed the dynamics of diffusion processes, Figure 2 highlights the news diffusion processes of two events: the birth of "Royal Baby" Prince George (circles) and the publication of a United Nations (UN) report on the use of chemical weapons in the civil war in Syria (triangles). The diffusion processes of both events are similar with regard to their full length: within the 24-hour time-span observed, it takes 13.7 hours for the last online news site to report the birth of the "Royal Baby" and 14.9 hours to cover the UN report on Syria. Obviously, the diffusion of both events differs in terms of range: after 24 hours, the number of online news sites reporting about the "Royal Baby" has grown to 26, while there have been 21 news sites covering the UN report. Moreover, from visual inspection of process dynamics, we can conclude that the diffusion of news about the "Royal Baby" approaches saturation many hours early relative to its full length. The number of online news sites reporting the childbirth accumulates at fast pace during a short time-span after the event has been mentioned first, while the rate of reporting-site accumulation slows down significantly for the remaining, longer period of the process. To account for the burst at the beginning of the diffusion process, we calculated the duration of the main diffusion phase, defined as the time-span between the first online news site reporting the event (just like for the full length of the process) and the point during the process when the rate of reporting-site accumulation decreases significantly. For the "Royal Baby", this point has been reached after 1.6 hours, when 23 online news sites had reported the event. In contrast, the diffusion of news about the UN report on Syria lacks a clear transition from a news flow characterized by surges and bursts at the beginning to a slower rate of accumulation during the remaining process. That is why we assumed its main diffusion phase would last during the full process, 14.9 hours. As our analysis aims to describe general temporal patterns of news diffusion processes beyond case studies, we will now provide summary statistics for range, length, and dynamics of all 95 event-based news diffusion processes.

The effective total sample size of $N = 1919$ articles corresponds to an average range of mean = 20.2 (SD = 5.1) of 28 online news sites reporting each event within 24 hours. Four events were reported by all 28 news sites: German political satire artist Dieter Hildebrandt's death in November 2013, Nelson Mandela's death in December 2013, German Chancellor Angela Merkel's skiing accident, and former professional soccer player Thomas Hitzelperger's coming-out (both in January 2014). Another four events were reported by 27 news sites, e.g., allegations of surveillance of Chancellor Merkel's mobile phone against the NSA, and German Green Party politician Hans-Christian Ströbele's meeting with Edward Snowden in Moscow (both in October 2013). Overall, the frequencies of the number of online news sites reporting

each event follow the pattern of a bimodal distribution with its low point at 17 online news sites. Accepting this low point to split the distribution, there were 27 lower-range news diffusion processes involving up to 17 online news sites (28 percent) and 68 higher-range diffusion processes involving more than 17 news sites (72 percent).

The average length of diffusion processes, indicated by the time lag of the last online news site reporting an event within 24 hours, was mean = 48,751 seconds (SD = 23,748 seconds), 13.5 hours. Similar to the diffusion of news about the "Royal Baby", many diffusion processes were characterized by initial surges and early saturation relative to their full length (see Figure 3). Across all 95 events, the vast majority of reports were published during the first 10 percent of relative length of diffusion processes (see Figure 4). On average, 7.5 online news sites have reported each event after 30 minutes, 10.6 sites after 60 minutes, and 12.6 sites after 120 minutes. Accordingly, main diffusion phases are significantly shorter than full diffusion processes, lasting mean = 18,219 seconds (SD = 20,091 seconds), 5.1 hours, and comprising mean = 17.8 reporting online news sites (SD = 4.9). The distribution of the length of main diffusion phases is strongly skewed to the right with its median as early as 7916 seconds, 2.2 hours.

Subsamples of Recurring Dynamic Patterns of Diffusion Processes

To get a more fine-grained idea of the dynamics of news flows among news sites, we turn to the cluster solution obtained for the time-lag data of each diffusion process. Based on their dynamics, we identified three subsamples consisting of $N = 43$ (45 percent; Cluster 1), $N = 28$ (29 percent; Cluster 2), and $N = 24$ (25 percent; Cluster 3) news diffusion processes. In Figure 3, diffusion processes belonging to the same cluster are grouped visually by line shape. To display generalized dynamic patterns of Clusters 1, 2, and 3, we

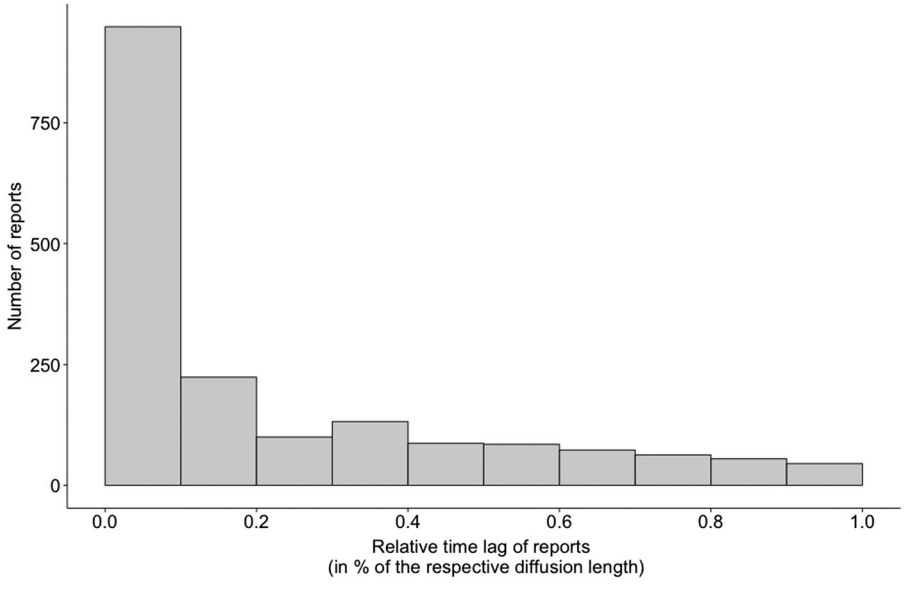


FIGURE 4
Number of first reports by online news sites during diffusion processes

aggregated the time-lag data of every diffusion process within each cluster to a single diffusion curve summarizing each cluster in Figure 5. As the data span across the whole subsample of diffusion processes within each cluster in this analysis, instead of absolute numbers, we display the proportion of online news sites reporting (from totally 28) in relation to the amount of time elapsed. Visual inspections of Figures 3 and 5 suggest primarily range and rate of reporting news sites accumulating distinguish between the clusters. In Cluster 1, 42 out of 43 diffusion processes (97.7 percent) belong to the subsample of higher-range diffusion processes comprising more than 17 news sites; in Cluster 2, this share is at 64.3 percent ($N = 18$); and in Cluster 3, it is at 33.3 percent ($N = 8$) ($\chi^2 = 32.382$; df = 2; $p < 0.001$). Diffusion processes in Cluster 1 (mean = 41,327 seconds, 11.5 hours; SD = 26,415 seconds) are similar to those in Cluster 2 (mean = 48,206 seconds, 13.4 hours; SD = 21,735 seconds) in terms of their full length; diffusion processes of both Clusters 1 and 2 are finished significantly earlier than processes of Cluster 3 (mean = 62,687 seconds, 17.4 hours; SD = 13,042 seconds) (Welch = 11.121; df = 2; $p < 0.001$).

Revealing the dynamics of diffusion rates, it takes only mean = 5545 seconds (SD = 3559 seconds), 1.5 hours, for processes in Cluster 1 to arrive at the point of saturation and therefore at the end of their main diffusion phases—significantly earlier than processes in Cluster 2 (mean = 18,413 seconds, 5.1 hours; SD = 15,420 seconds). Processes in Cluster 3 are the slowest to arrive at this point (mean = 40,702 seconds, 11.3 hours; SD = 22,795 seconds) (Welch = 36,212; df = 2; $p < 0.001$). In addition to the temporal lead of processes in Cluster 1 at the end of the main diffusion phases, the number of online news sites reporting has accumulated to mean = 20.8 (SD = 3.2) by then, which is substantially higher than for diffusion processes in Cluster 2 (mean = 16.3; SD = 4.5) and in Cluster 3 (mean = 14.3; SD = 5.0) (Welch = 21.802; df = 2; $p < 0.001$). Hence, the main diffusion phases of processes in

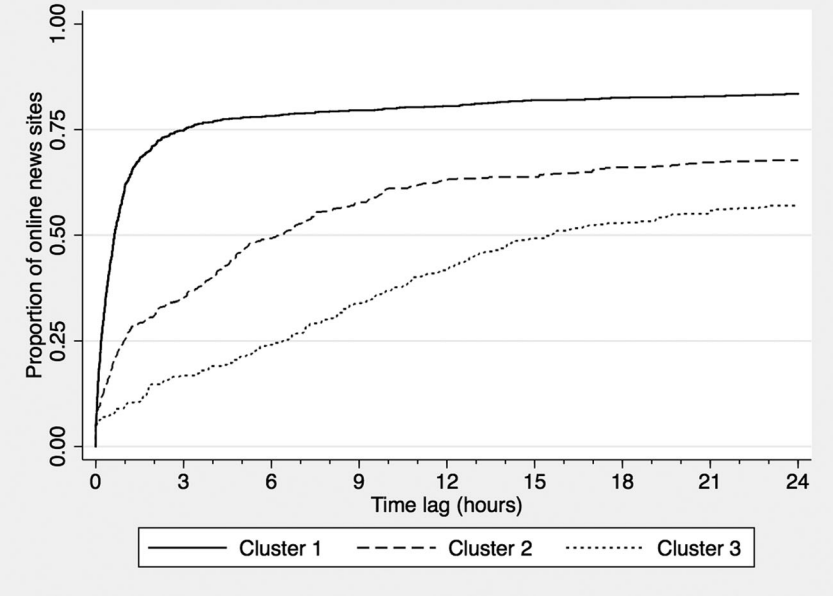


FIGURE 5
Diffusion processes aggregated by dynamics-based cluster membership

Cluster 1 are not only substantially shorter than those of processes in Clusters 2 and 3, but they also comprise higher numbers of online news sites (see Figures 3 and 5).

The combination of both attributes constitutes process dynamics characterized by surges and bursts in Cluster 1 shortly after first reports about the respective events. For example, 14 of 28 online news sites have reported the birth of "Royal Baby" Prince George within 30 minutes, 21 within 60 minutes, and 23 within 120 minutes. On average, for each process in Cluster 1, the number of online news sites reporting has accumulated to 12.2 after 30 minutes, to 17.2 after 60 minutes, and to 20.0 after 120 minutes (see also proportions across all events in each cluster in Figure 5). Typically, events diffusing among online news sites at this fast pace do not afford the addition of contextual information to be unambiguously grasped by the audience and thus suit instantaneous coverage. Often they are characterized by negative outcomes. From the events sampled, 11 of 12 accidents or deaths involving famous personalities result in diffusion processes exhibiting surges in news flow in Cluster 1, e.g., former Israeli Prime Minister Ariel Sharon's death, Nelson Mandela's death, German Chancellor Merkel's skiing accident, and Michael Schumacher's skiing accident. Additionally, five of eight accidents or disasters with many people involved result in diffusion processes belonging to Cluster 1, e.g., two train accidents in France and Spain, respectively. But we also find 9 of 13 court decisions or voting results among the news diffusion processes in Cluster 1, e.g., a jail sentence against suspected murderer Amanda Knox, the exclusion of former Prime Minister Silvio Berlusconi from the Italian Senate, and Munich citizens' referendum against the city's application for the Winter Olympics—that is, events that can be anticipated.

The surges in news diffusion processes shortly after the very first report about an event is less distinct for events in Cluster 2. Similar to Cluster 1, diffusion processes in Cluster 2 approach saturation early in comparison to their full length. However, both in total and relative to the amount of time elapsed, they do not accumulate as many reporting online news sites as processes in Cluster 1. On average, during diffusion processes in Cluster 2, 4.9 news sites have reported an event after 30 minutes, 7.1 after 60 minutes, and 8.7 after 120 minutes (see also proportions in Figure 5). The sample of events resulting in diffusion processes in Cluster 2 is diverse, comprising definitely relevant, but—from the German point of view—non-extraordinary news from a variety of fields like politics, economy, and people, e.g., the release of Russian political activist Alexei Navalny from prison, Amazon founder Jeff Bezos' acquisition of the *Washington Post*, and Swedish Princess Madelaine's pregnancy.

For events in Cluster 3, there is much higher correspondence between the endpoints of main diffusion phases and the (preliminary) endpoints of full diffusion processes. The rate of reporting news sites accumulating tends to be constant instead of evolving from faster to slower, saturated phases. Diffusion processes in Cluster 3 rather unfold slowly during relatively longer periods of time. On average, they have involved 2.1 online news sites after 30 minutes, 2.7 after 60 minutes, and 4.1 after 120 minutes (see also proportions in Figure 5). Additionally, their range is typically less extensive than for processes in Clusters 1 and 2. The diffusion of news about the UN report on chemical weapons in Syria belongs to Cluster 3, for example, as well as economy news without direct effects on German news consumers, e.g., the withdrawal of agricultural corporation Monsanto from European markets and the bankruptcy of Detroit. Oftentimes, these stories require additional investigation of contextual information in the newsroom to be made fully comprehensible to German readers.

Conclusion and Discussion

In this study, we observed the dynamics of 95 event-based news diffusion processes among 28 German online news sites. We identified three recurring temporal patterns of news diffusion processes, which systematically differed in terms of their range and the pace at which online news sites joined the diffusion process. Almost half of the diffusion processes exhibit a temporal pattern of abrupt bursts shortly after events have been reported for the very first time; in these cases, the timing of publication decisions among the various newsrooms in the news-site ecosystem is extraordinarily narrow. These diffusion processes often reach saturation within 1.5 hours only, and involve a majority of online news sites. The rapid dynamics of these diffusion processes suggest that many online newsrooms share the eagerness to instantaneously report respective events such as the birth of "Royal Baby" Prince George. Often, these news stories were easily comprehensible and characterized by negative outcomes. Yet we also found less spectacular, but anticipated events diffuse by the same dynamic pattern, e.g., the judgments against Amanda Knox and Silvio Berlusconi as well as the Munich referendum against applying for the Winter Olympics. This observation supports the notion that instantaneous news coverage has been expanded beyond dramatic or at least extraordinary events (Rosenberg and Feldman 2008; Lewis and Cushion 2009) and is a common characteristic of working practices in online newsrooms today. Here, the opportunity to keep the audience up-to-date about major turning points of established story lines seems to be the driver of instantaneous event coverage.

However, instantaneous coverage as response strategy does not appear to apply equally to every event for each online news site. We also observed news diffusion processes for which the number of reporting news sites accumulated at much slower rates during longer time-frames, e.g., the publication of an UN report on chemical weapons in Syria, which confirmed allegations raised earlier in the conflict. Respective events are not characterized by the exceptionally unambiguous relevance of "news events", for which "breaking news" coverage appears as a common working practice within the whole news-site ecosystem. Rather, prompt publishing responses to events are part of the working routines of some online newsrooms here as well, while a variety of other online news sites in the ecosystem take more time to disseminate these stories to their audiences.

Analytically complementing empirical accounts of the evolvement of stories by single online news providers with an ecosystem perspective, the present study contributes to insights into how the acceleration of news cycles shapes working practices in online journalism. Focusing on the timing of event-based publication decisions by 28 German online newsrooms, the results reveal that the prompt release of news to the public is a practice widely shared within the news-site ecosystem. This finding underlines the prevalence of technological opportunities and working routines fostering timely publishing responses to events within the whole news-site ecosystem. In consequence, news about the latest events is frequently made ubiquitously available to the public during very short time-frames, often within the first 1.5 hours, no matter which news provider individual media users turn to. However, the results also demonstrate that brief news diffusion processes with initial bursts do not characterize the diffusion of all types of news equally. Compared to "news events" and major turning points of established story lines, when the relevance of events seems less clear for the news sites' audiences, the time-span of their article publications widens. A likely explanation for this distinction between the dynamics of news

diffusion processes is that the pursuit of immediacy as a norm of online news production is widely shared in the news-site ecosystem for "news events" and story-line turning points, while, for other events, there is variation of the pursuit of immediacy among newsrooms.

Methodologically, a main purpose of this study was to explore the potentials of a machine-based content analysis to systematically describe news diffusion processes among online news sites based on a large-scale database. In doing so, we proposed a methodological alternative to studies of online communication processes that depend on the analysis of direct references between documents in the form of source mentions or hyperlinks within a text. In our analysis, pace and dynamics of the reconstructed diffusion processes are established by retracing the timing of multiple online news sites covering the same real-life events. This way, we were able to uncover the adoption of news within the ecosystem beyond direct imitations among online news providers. By focusing on the output of the news process, especially on the question of how quickly news is made available to the public by online news providers, our event-based approach adds a new perspective to the study of online news flows.

Limitations of the study originate from sampling procedures and the volatility of online news content. The sample was restricted to events characterized by high newsworthiness, thus necessarily leaving questions about relations between variations of newsworthiness and patterns of news diffusion processes unanswered. Although we made use of virtually real-time content data collection, we sometimes faced the problem that articles published at a specific point in time had later been updated (Karlsson and Strömbäck 2010). We tried to handle this limitation to the accuracy of publication-time data by screening the contents of potentially problematic articles as well as surrounding articles. Likewise, we had to manually scan the content of articles when searching the database with keywords did not unambiguously identify events, either because online news sites announced predictable events before they were actually happening, or because there were several (semantically) similar events happening on the same day (e.g., Russia's amnesties for members of the punk band Pussy Riot and for former industrialist Mikhail Khodorkovsky on the same day in December 2013).

Despite these challenges to extracting event-based data from online news, this paper introduced the basic analytical tools to further explore the dynamics of news diffusion processes online. The research strategy has proven to be a fruitful approach to generate more detailed insights into commonalities and distinctions among the dynamics of online news flows. To better understand these processes, we aim to follow up on this work by comparing our results with news diffusion processes from various subdomains of the online news ecosystem, e.g., from a variety of countries. Below the level of the macro-dynamics of diffusion processes, future studies could aim at explaining the underlying mechanisms for the emergence of news-flow dynamics. For example, our results indicate that individual online news sites regularly report in a similar position within the diffusion processes, with national and regional news hubs (*Spiegel*, *Kölner Stadtanzeiger*, *Welt*, *Mitteldeutsche Zeitung*) being regularly among the first to report events. These micro-dynamics require further exploration. Additionally, future studies should aim to include news agencies when reconstructing the news ecosystem, as they regularly play a crucial role in breaking news online by providing a variety of online newsrooms with the same stories at the same time. Likewise, social media are an important source of information for journalists, e.g., when citizen reporters coincidentally witness a newsworthy event. Information provided through these channels could be tracked in future research, too, to gain a more comprehensive map of the online

news ecosystem. At the story level, considering the distinctions between news diffusion processes dependent on event type, it seems likely that newsworthiness partly determines news diffusion dynamics. However, it is unclear whether it is a simple addition of news factors, or whether it is the presence of specific news factors that causes, for example, bursts of diffusion curves. Finally, tracing follow-up story coverage would combine the ecosystem perspective of the present study with a story-evolvement approach.

In this paper, we provided empirical data making the potentials for brevity and abrupt surges in news diffusion processes within the online news ecosystem observable. Relying on a broad sample of events instead of case studies only, we were able to reconstruct the recurrence of temporal patterns of news diffusion processes among online news sites. As this is, to the best of our knowledge, the first study that empirically reconstructs online news diffusion processes at the ecosystem level, this study has also raised a lot of new questions. With the help of innovative research methods (Günther and Quandt 2016), we can deepen our understanding of the velocity of online news beyond the individual online newsroom, and shed new light onto the flow of online news.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

This work was partially supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) [grant number 01UG1232A].

REFERENCES

Anderson, Christopher W. 2010. "Journalistic Networks and the Diffusion of Local News: The Brief, Happy News Life of the "Francisville Four"." *Political Communication* 27 (3): 289–309. doi:10.1080/10584609.2010.496710.

Benkler, Yochai. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press.

Berkowitz, Dan. 1992. "Non-Routine News and Newswork: Exploring a What-A-Story." *Journal of Communication* 42 (1): 82–94. doi:10.1111/j.1460-2466.1992.tb00770.x.

Boczkowski, Pablo J. 2010. *News at Work: Imitation in an Age of Information Abundance*. Chicago: University of Chicago Press.

Boczkowski, Pablo J., and Martin de Santos. 2007. "When More Media Equals Less News: Patterns of Content Homogenization in Argentina's Leading Print and Online Newspapers." *Political Communication* 24 (2): 167–180. doi:10.1080/10584600701313025.

De Fleur, Melvin L. 1987. "The Growth and Decline of Research on the Diffusion of the News, 1945–1985." *Communication Research* 14 (1): 109–130. doi:10.1177/009365087014001006.

Donsbach, Wolfgang. 2004. "Psychology of News Decisions: Factors behind Journalists' Professional Behavior." *Journalism* 5 (2): 131–157. doi:10.1177/146488490452002.

Galtung, Johan, and Mari Holmboe Ruge. 1965. "The Structure of Foreign News: The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers." *Journal of Peace Research* 2 (1): 64–90. doi: 10.1177/002234336500200104.

Glascock, Jack, and Larry J. King. 2007. "The Columbia Shuttle Breakup: News Diffusion Comparison of a Local and Distant Location." *Communication Research Reports* 24 (3): 205–213. doi:10.1080/08824090701439059.

Greenberg, Bradley S. 1964. "Diffusion of News of the Kennedy Assassination." *Public Opinion Quarterly* 28 (2): 225–232. doi: 10.1086/267239.

Günther, Elisabeth, and Thorsten Quandt. 2016. "Word Counts and Topic Models: An Overview of Automated Text Analysis Methods for Digital Journalism Research." *Digital Journalism* 4 (1): 75–88. doi:10.1080/21670811.2015.1093270.

Heinrich, Ansgard. 2012. "Foreign Reporting in the Sphere of Network Journalism." *Journalism Practice* 6 (5/6): 766–775. doi:10.1080/17512786.2012.667280.

Hermida, Alfred. 2010. "Twittering the News: The Emergence of Ambient Journalism." *Journalism Practice* 4 (3): 297–308. doi:10.1080/17512781003640703.

Im, Yung-Ho, Eun-mee Kim, Kyungmo Kim, and Yeran Kim. 2011. "The Emerging Mediascape, Same Old Theories? A Case Study of Online News Diffusion in Korea." *New Media & Society* 13 (4): 605–625. doi:10.1177/1461444810377916.

Karlsson, Michael, and Jesper Strömbäck. 2010. "Freezing the Flow of Online News: Exploring Approaches to the Study of the Liquidity of Online News." *Journalism Studies* 11 (1): 2–19. doi:10.1080/14616700903119784.

Kohlschütter, Christian, Peter Fankhauser, and Wolfgang Nejdl. 2010. "Boilerplate Detection Using Shallow Text Features." Proceedings of the Third ACM International Conference on Web Search and Data Mining. New York: ACM, 441–50.

Lewis, Justin, and Stephen Cushion. 2009. "The Thirst to Be First: An Analysis of Breaking News Stories and Their Impact on the Quality of 24-Hour News Coverage in the UK." *Journalism Practice* 3 (3): 304–318. doi:10.1080/17512780902798737.

Lewis, Justin, Stephen Cushion, and James Thomas. 2005. "Immediacy, Convenience or Engagement? An Analysis of 24-Hour News Channels in the UK." *Journalism Studies* 6 (4): 461–477. doi:10.1080/14616700500250362.

Lim, Jeongsub. 2011. "Intermedia Agenda Setting and News Discourse: A Strategic Response Model for a Competitor's Breaking Stories." *Journalism Practice* 5 (2): 227–244. doi:10.1080/17512786.2010.509184.

Lim, Jeongsub. 2012. "The Mythological Status of the Immediacy of the Most Important Online News: An Analysis of Top News Flows in Diverse Online Media." *Journalism Studies* 13 (1): 71–89. doi:10.1080/1461670X.2011.605596.

Messner, Marcus, and Marcia Watson DiStaso. 2008. "The Source Cycle: How Traditional Media and Weblogs Use Each Other as Sources." *Journalism Studies* 9 (3): 447–463. doi:10.1080/14616700801999287.

Quandt, Thorsten, Martin Löffelholz, David H. Weaver, Thomas Hanitzsch, and Klaus-Dieter Altmeppen. 2006. "American and German Online Journalists at the Beginning of the 21st Century: A Bi-National Survey." *Journalism Studies* 7 (2): 171–186. doi:10.1080/14616700500533486.

Reese, Stephen D., and Lucig H. Danielian. 1989. "Intermedia Influence and the Drug Issue: Converging on Cocaine." In *Communication Campaigns About Drugs: Government, Media, and the Public*, edited by Pamela J. Shoemaker, 29–45. Mahwah: Lawrence Erlbaum.

Risley, Ford. 2000. "Newspapers and Timeliness: The Impact of the Telegraph and the Internet." *American Journalism* 17 (4): 97–103. doi:10.1080/08821127.2000.10739275.

Rogers, Everett M. 2000. "Reflections on News Event Diffusion Research." *Journalism & Mass Communication Quarterly* 77 (3): 561–576. doi:10.1177/107769900007700307.

Rogers, Everett M., and Nancy Seidel. 2002. "Diffusion of News of the Terrorist Attacks of September 11, 2001." *Prometheus* 20 (3): 209–219. doi:10.1080/0810902021014326.

Rosenberg, Howard, and Charles S. Feldman. 2008. *No Time to Think: The Menace of Media Speed and the 24-Hour News Cycle*. New York: Continuum International.

Saltzis, Kostas. 2012. "Breaking News Online: How News Stories Are Updated and Maintained Around-the-Clock." *Journalism Practice* 6 (5/6): 702–710. doi:10.1080/17512786.2012.667274.

Schütz, Walter J. 2012. Deutsche Tagespresse 2012: Ergebnisse der Aktuellen Stichttagssammlung [German Daily Press 2012: Results of the Recent Key Data]. *Media Perspektiven* 11: 570–593. http://www.media-perspektiven.de/publikationen/fachzeitschrift/2012/artikel/deutsche-tagespresse-2012/.

Singer, Jane B. 2003. "Who Are These Guys? The Online Challenge to the Notion of Journalistic Professionalism." *Journalism* 4 (2): 139–163. doi:10.1177/146488490342001.

Spitzberg, Brian H. 2014. "Toward A Model of Meme Diffusion (M^3D)." *Communication Theory* 24 (3): 311–339. doi:10.1111/comt.12042.

Vliegenthart, Rens, and Stefaan Walgrave. 2008. "The Contingency of Intermedia Agenda Setting: A Longitudinal Study in Belgium." *Journalism & Mass Communication Quarterly* 85 (4): 860–877. doi:10.1177/107769900808500409.

Weber, Matthew S., and Peter Monge. 2011. "The Flow of Digital News in a Network of Sources, Authorities, and Hubs." *Journal of Communication* 61 (6): 1062–1081. doi:10.1111/j.1460-2466.2011.01596.x.

Widholm, Andreas. 2016. "Tracing Online News in Motion: Time and Duration in the Study of Liquid Journalism." *Digital Journalism* 4 (1): 24–40. doi:10.1080/21670811.2015.1096611.

Florian Buhl (author to whom correspondence should be addressed), Department of Communication, University of Münster, Germany. E-mail: florian.buhl@uni-muenster.de. Web: http://www.uni-muenster.de/Kowi/personen/florian-buhl.html ORCID http://orcid.org/0000-0002-4629-6869

Elisabeth Günther, Department of Communication, University of Münster, Germany. E-mail: elisabeth.guenther@uni-muenster.de. Web: http://www.uni-muenster.de/Kowi/personen/elisabeth-guenther.html ORCID http://orcid.org/0000-0001-5754-973X

Thorsten Quandt, Department of Communication, University of Münster, Germany. E-mail: thorsten.quandt@uni-muenster.de. Web: http://www.uni-muenster.de/Kowi/personen/thorsten-quandt.html ORCID http://orcid.org/0000-0003-1937-0282

Appendix A

Keywords Used to Identify Event-related Online News Reports

			Keywords			
	Date	**Description**	**Mandatory1**	**Mandatory2**	**Optional1**	**Optional2**
1	07/12/13	Edward Snowden applies for asylum in Russia	Snowden		Asyl [asylum]	Russland [Russia]
2	07/12/13	Boeing 787 catches fire at London Heathrow airport	Boeing		London	Heathrow
3	07/12/13	Paris train crash	Zugunglück [train crash]		Paris	Frankreich [France]
4	07/14/13	Jury acquits Zimmerman of all charges in Trayvon Martin case	Freispruch [acquittal]		Trayvon	Zimmerman
5	07/15/13	German BND cooperates with NSA	BND [foreign intelligence agency]		NSA	
6	07/17/13	Decision to host the World Cup in Qatar in winter	WM [world cup]		Katar [Qatar]	Winter [winter]
7	07/18/13	Monsanto will halt production of genetically modified corn in Europe	Monsanto		Gentechnik [genetic engineering]	Gentechnisch [gene-modified]
8	07/19/13	Detroit, Michigan, files for bankruptcy	Detroit		Insolvent [insolvent]	Insolvenz [insolvency]
9	07/19/13	Kremlin-critic Navalny freed on bail	Nawalny			
10	07/19/13	Famous goalkeeper Trautmann dies aged 89	Trautmann			
11	07/22/13	Royal baby announcement	William		Geburt [birth]	Sohn [son]
12	07/23/13	O2 acquires E-Plus	O2			
13	07/24/13	Google reveals Chromecast	Chromecast			
14	07/25/13	Spain train crash	Zugunglück [train crash]		Spanien [Spain]	Santiago
15	07/25/13	German DIY store Max Bahr insolvent	Bahr	Max	Insolvenz [insolvency]	Insolvent [insolvent]

16	07/26/13	Arrest warrant issued for Morsi	Mursi [Morsi]		Haftbefehl [arrest warrant]	Staatsanwaltschaft [prosecution]
17	07/29/13	Accusation of plagiarism against Lammert, President of German Parliament	Lammert	Norbert	Plagiat [plagiarism]	Doktorarbeit [PhD thesis]
18	08/02/13	Russia grants Snowden asylum	Snowden		Asyl [asylum]	Russland [Russia]
19	08/02/13	Mysterious mummy found in German attic	Mumie [mummy]			
20	08/05/13	Amazon boss buys *Washington Post*	Post	Washington	Bezos	
21	08/12/13	Dutch Prince Friso dies after coma	Friso			
22	08/13/13	German leftist politician Bisky dies	Bisky			
23	08/19/13	Man holds hostages at Ingolstadt town hall	Ingolstadt		Geiseln [hostages]	Geiselnahme [hostage-taking]
24	08/21/13	Syrian government accused of using lethal gas in attacks	Syrien [Syria]		Giftgas [lethal gas]	Giftgasangriff [chemical attack]
25	08/22/13	German tourist dies after Hawaii shark attack	Hawaii		Hai [shark]	Haiangriff [shark attack]
26	08/23/13	Microsoft CEO Ballmer resigns	Ballmer		Microsoft	
27	09/01/13	Vodafone sells Verizon	Vodafone		Verizon	
28	09/01/13	Morsi's impeachment	Mursi [Morsi]		Anklage [impeachment]	Staatsanwaltschaft [prosecution]
29	09/03/13	Microsoft acquires Nokia mobile business	Microsoft		Nokia	
30	09/03/13	Swedish Princess Madeleine announces pregnancy	Madeleine		Schwanger [pregnant]	Schwangerschaft [pregnancy]
31	09/07/13	German Democrat opponent Steinbrück blackmailed	Steinbrück		Putzfrau [cleaning lady]	Erpressung [blackmail]
32	09/11/13	German court rules Muslim girls must join swimming class	Schwimmunterricht [swimming class]		Muslimin [Muslim]	Islam
33	09/12/13	Vodafone Germany confirms data theft	Vodafone		Daten [data]	Kunden [customers]
34	09/12/13	German actor Otto Sander dies	Sander	Otto		

(*Continued*)

Appendix A

(*Continued*)

			Keywords			
	Date	**Description**	**Mandatory1**	**Mandatory2**	**Optional1**	**Optional2**
35	09/16/13	UN report on chemical weapons use in Syria	UN		Giftgas [lethal gas]	Giftgaseinsatz [chemical attacks]
36	09/16/13	Pedophilia scandal related to German Green Party leader Trittin	Trittin		Pädophilie [pedophilia]	Göttingen
37	09/22/13	Hamburg citizens vote to buy back energy grid	Hamburg		Volksentscheid [referendum]	Vattenfall
38	09/25/13	Ted Cruz's marathon speech against Obamacare	Cruz	Ted		
39	09/27/13	Social Democrats ask party members to decide on joining Merkel coalition	SPD		Basis [basis]	Mitgliederentscheid [member's vote]
40	10/04/13	Refugee tragedy in Lampedusa	Lampedusa		Flüchtling [refugee]	Flüchtlinge [refugees]
41	10/14/13	Ireland leaves European bailout fund	Irland [Ireland]		Rettungsschirm [bailout fund]	Eurozone [euro zone]
42	10/14/13	US bankruptcy shutdown	USA		Shutdown	Haushalt [budget]
43	10/15/13	Yasser Arafat exhumed	Arafat		Vergiftung [poisoning]	Vergiftet [poisoned]
44	10/17/13	EU court rules digital biometric data in passports are legal	Reisepass [passport]		Fingerabdruck [finger print]	Fingerabdrücke [finger prints]
45	10/19/13	Italian court rules Berlusconi is barred from serving any legislative office	Berlusconi		Urteil [sentence]	Berufungsgericht [appellate court]
46	10/23/13	Alleged US surveillance of Merkel's phone	Merkel		Handy [mobile phone]	Mobiltelefon [mobile phone]
47	10/31/13	German politician Ströbele meets with Snowden	Ströbele		Snowden	
48	11/03/13	Munich authorities discover 1500 works of art stolen by Nazis	München [Munich]		Kunstraub [art theft]	Kunstwerke [works of art]

49	11/04/13	Bayern Munich president Hoeneß accused of tax fraud	Hoeneß		Steuerhinterziehung [tax fraud]	Anklage [impeachment]
50	11/10/13	Referendum against Olympic Games in Munich	München [Munich]		Olympia [Olympic Games]	Winterspiele [Winter Olympics]
51	11/11/13	Typhoon Haiyan hits Philippines	Philippinen [Philippines]		Taifun [typhoon]	Haiyan
52	11/14/13	Re-election of Sigmar Gabriel as SPD chairman	Gabriel		Vorsitzender [chairman]	Parteivorsitzender [party chairman]
53	11/20/13	Cabaret artist Dieter Hildebrandt dies	Hildebrandt	Dieter		
54	11/27/13	Berlusconi loses Italian Senate seat	Berlusconi		Senat [senate]	Senatssitz [senate seat]
55	11/29/13	Interview between SPD chairman Gabriel and TV host Slomka	Slomka	Marietta	Gabriel	
56	11/30/13	Glasgow helicopter crash	Glasgow		Hubschrauber [helicopter]	Helikopter [helicopter]
57	12/02/13	Amazon plans drone delivery	Amazon		Drohne [drone]	Drohnen [drones]
58	12/06/13	Nelson Mandela dies	Mandela	Nelson		
59	12/08/13	German President Gauck boycotts Sochi Winter Olympics	Gauck		Sotschi [Sochi]	
60	12/11/13	Uruguay legalizes cannabis	Uruguay		Cannabis	Marihuana
61	12/19/13	Putin announces amnesty for punk band Pussy Riot	Riot	Pussy	Amnestie [amnesty]	
62	12/19/13	Putin announces amnesty for Khodorkovsky	Chodorkowski [Khodorkovsky]			
63	12/29/13	Formula 1 legend Michael Schumacher's skiing accident	Schumacher	Michael		
64	01/05/14	Soccer legend Eusébio dies	Eusebio			
65	01/06/14	Merkel's skiing accident	Merkel		Beckenring [pelvis]	Langlauf [cross-country skiing]
66	01/08/14	German soccer player Hitzlsperger has coming-out	Hitzlsperger		Homosexualität [homosexuality]	Homosexuell [homosexual]
67	01/11/14	Ariel Sharon dies	Scharon [Sharon]	Ariel		
68	01/11/14	Wind power company Prokon is insolvent	Prokon		Insolvenz [insolvency]	Insolvent [insolvent]

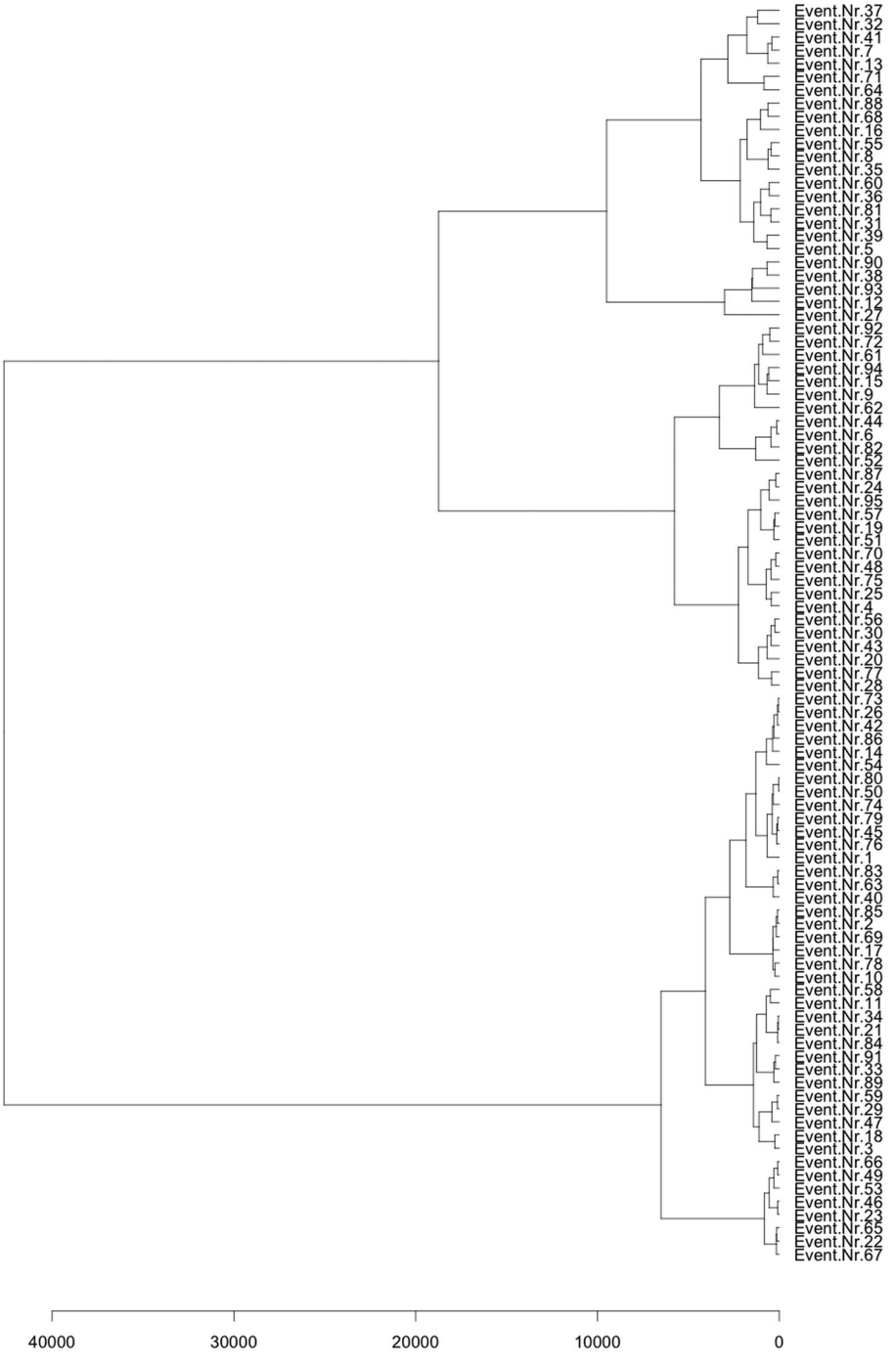
(*Continued*)

Appendix A

(*Continued*)

			Keywords			
	Date	**Description**	**Mandatory1**	**Mandatory2**	**Optional1**	**Optional2**
69	01/13/14	Ritter-Sport wins lawsuit against Stiftung Warentest	Warentest	Stiftung	Ritter	
70	01/14/14	Automobile club ADAC fabricates results of "favorite car" award	ADAC		Lieblingsauto [favorite car]	
71	01/26/14	Women and children are allowed to leave Homs	Homs		Brahimi	Friedenskonferenz [peace conference]
72	01/29/14	Google sells Motorola to Lenovo	Google		Motorola	Lenovo
73	01/30/14	Germany's Federal Cartel Office initiates proceedings against rail incumbent Deutsche Bahn	Bahn	Deutsche	Bundeskartellamt [Federal Cartel Office]	Kartellamt [Cartel Office]
74	01/30/14	Amanda Knox is sentenced to 28 years in prison	Knox	Amanda		
75	02/02/14	Woody Allen accused of sexual abuse	Allen	Woody	Missbrauch [abuse]	Missbraucht [abused]
76	02/02/14	Philip Seymour Hoffman dies	Seymour	Philip	Hoffman	
77	02/06/14	US state department's diplomat for Europe Nuland says "Fuck the EU"	EU	Fuck	Nuland	
78	02/09/14	Swiss voters back referendum to bring back quotas for immigration	Mehrheit [majority]		Volksabstimmung [referendum]	Masseneinwanderung [mass immigration]
79	02/11/14	Shirley Temple dies	Temple	Shirley		
80	02/11/14	More artworks found in Gurlitt's house in Salzburg	Gurlitt		Salzburg	Salzburger
81	02/13/14	Senator Rand Paul sues Obama over NSA spying	Paul	Rand	Obama	
82	02/13/14	Italian Prime Minister Enrico Letta resigns	Letta	Enrico	Rücktritt [resignation]	

83	02/17/14	Hijacked Ethiopian Airlines plane lands in Geneva	Airlines	Ethiopian	Genf [Geneva]	
84	02/19/14	Facebook acquires messaging service WhatsApp	Facebook		Whatsapp	
85	02/20/14	Deutsche Bank cuts deal in Kirch case	Vergleich [deal]		Kirch	
86	02/22/14	Drug lord "El Chapo" Guzman caught in Mexico	Chapo	El		
87	02/24/14	Edathy case: charges against prosecutors	Dienstgeheimnis [official secret]	Verletzung [violation]	Edathy	
88	02/25/14	Far-right National Democratic Party of Germany (NPD) sues President Gauck	Gauck		Wieland	
89	02/26/14	Germany's Federal Constitutional Court scraps 3 percent vote threshold for EU poll	Bundesverfassungsgericht [Germany's Federal Constitutional Court]		Sperrklausel [barring clause]	Dreiprozent-Hürde [three-percent rule]
90	02/27/14	Judge rules same-sex marriage is legal in Texas	Ehe [marriage]	Homo [same-sex]	Texas	
91	02/27/14	Germany's former President Wulff acquitted of all charges	Wulff	Christian	Freispruch [acquittal]	Freigesprochen [acquitted]
92	02/28/14	Catholic cardinal Meisner resigns	Meisner	Kardinal [cardinal]	Ruhestand [retirement]	Rücktrittsgesuch [letter of resignation]
93	03/06/14	Leader of a Catholic workers' lobbying group Hupfauer resigns after child porn allegations	Hupfauer		Rücktritt [resignment]	Zurückgetreten [resigns]
94	03/06/14	Russian politicians face sanctions by US for Crimea Crisis	Russland [Russia]		Einreiseverbote [entry bans]	Einreiseverboten [entry bans]
95	03/09/14	Missing Malaysia Airlines airplane MH370	Airlines	Malaysia	Boeing	

Appendix B

Events with Similar Dynamics of News Diffusion Processes (Ward Hierarchical Clustering)